Rumsey, Michael G., US Army Research Institute for the Behavioral and Social Sciences, Alexandria, Virginia. (Thurs. P.M.)

Race Influences on Peer Ratings in ROTC Training Platoons

This study examined the influences of rater and ratee race on peer ratings of 4604 white and 884 black cadets distributed between three regional Army ROTC training camps. Blacks and whites were found to each give consistently higher ratings to their own subgroup than to the other, a tendency which was exacerbated when the minority subgroup judgments were particularly discrepant from the platoon judgments. The possible applicability of the concept of "race-bounded" friendships to these findings is considered. The pattern of black-white differences on peer ratings paralleled the pattern of such differences on other Advanced Camp measures.

RACE INFLUENCES ON PEER EVALUATIONS IN ARMY ROTC TRAINING PLATOONS

Michael G. Rumsey

US Army Research Institute
for the Behavioral and Social Sciences

Two recent literature reviews have presented a strong case for the utility of peer assessments. In their review, Lewin and Zwany (1976, p. 430) concluded: "In summary, peer evaluations are valid tools for predicting future success and are superior to all other measures available at the time of rating." Kane and Lawler (1978), reviewing a long list of studies reflecting positively on the validity of peer evaluations, suggested that the time for widespread operational use of this technique in work settings is at hand.

The growing enthusiasm for peer evaluations is understandable and, in a sense, overdue. Research on such evaluations has been conducted for many years, particularly in the military, with generally favorable results. Peer evaluation has been shown to have validity for predicting a variety of criteria, including future officer performance (Haggerty, 1963) and promotion (Downey, Medland, & Yates, 1976).

There are a number of reasons which may account for the favorable results associated with peer assessments. While other types of evaluators may be exposed to a limited sample of an individual's behavior, peers have the opportunity to observe performance in a wide variety of situations. Generally, peer evaluations also offer the opportunity to pool the observations of a substantial number of raters, an opportunity not inherent in other types of evaluations. This pooling has two fortuitous consequences: first, it operates to expand the behavioral base for the final rating and secondly, it reduces the impact of idiosyncratic rating tendencies associated with any given evaluator.

Kane and Lawler (1978, p. 555) suggested two reasons why, despite the empirical support for peer evaluation, its use in operational evaluation systems has historically been rather limited. One is the apparent confusion between peer assessment and sociometry, a procedure emphasizing personal preferences rather than evaluations of performance. The other is "the failure to recognize the need for its use." Were these the only bases for reluctance by managers to use a peer evaluation system, we might indeed expect that such reluctance would soon give way. However, I would suggest that there are a number of additional reasons, many of them associated with factors which could potentially compromise the accuracy of peer assessments. Peers typically have minimal training and experience as evaluators and may well have a personal involvement in the outcome of the evaluation process. Thus, they may be more susceptible than other evaluators to the influences of assessee characteristics which are fundamentally irrelevant to the evaluator's task. A companion paper (Rumsey, 1981) has examined the influence of one such characteristic, assessee gender; the present paper examines the influence of another, assessee race.

A number of studies (Cox & Krumboltz, 1958; DeJung & Kaplan, 1962; Mohr & Reidy, 1976) focusing on peer evaluations in predominantly white military units have provided evidence of same-race favoritism in the expression of these evaluations. Blacks tended to demonstrate a higher level of such favoritism than whites, but this finding may well be attributed to the blacks' minority status within the groups studied. DeJung and Kaplan (1962, p. 373) suggested as much, noting that the minority subgroup member, in rating others within this subgroup, might well be rating his or her "closest buddies." The majority member, in rating others in the majority subgroup, might also be rating his or her "closest buddies" but would be rating almost everyone else in the unit as well. Thus, the same-subgroup ratings given by majority members would not appear as consistently high as those given by minority members. This explanation assumes the existence of "race-bounded" friendships, a term used by Cox and Krumboltz (1958) as well as DeJung and Kaplan. We will return to a discussion of such friendships a little later.

In a study conducted by Schmidt and Johnson (1973) which examined groups composed of equal numbers of blacks and whites, no same-race favoritism was observed. Since these investigators exposed their subjects to human relations training, it was not possible to determine whether this training or the numerical racial equality was more responsible for the unbiased ratings in this study. However, another study conducted by Clore, Bray, Itkin and Murphy (1978) indicated that numerical equality may indeed be a significant factor. Here, after equal numbers of black and white children had attended summer camp together, both blacks and whites showed positive changes in their attitudes toward each other. The authors suggested that the numerical equality among black and white children, counselors and administrative staff contributed to an environment which eliminated status differences associated with race and thereby fostered the attitudinal changes which took place.

The present study examined race influences on peer nominations in the same environment as that studied by Mohr and Reidy (1976). As in this earlier study, Army ROTC cadets participated in a six week training camp, called Advanced Camp, and were organized into predominantly white platoons which typically contained a substantial minority of blacks. In the year between the Advanced Camp which was the focus of the previous study and the Advanced Camp which provided the data for the present study, however, the instructions for the peer nominations had been changed. In order to shift the focus of evaluation from personal feelings to performance, instructions were rewritten to emphasize "effectiveness" and "demonstrated contributions," whereas previously cadets were asked who they would be most and least willing to serve under. One purpose of the present study was to determine whether, given these new instructions, the previous finding of same-race favoritism would be replicated.

If such favoritism were found, this finding would activate a second purpose of this study: to examine the relationship between such favoritism and overall group disagreement. When judgments of a particular racial subgroup diverge from those of the overall group, they may do so because of same-subgroup preference or for a variety of other reasons. An exploration of whether same-subgroup preference was a major contributor to such disagreement was planned as a means of obtaining further understanding of racial influences on peer evaluations.

A final purpose of this study was to examine how peer evaluations received by blacks and whites compared with scores these subgroups received on other measures of Advanced Camp performance. Fortunately, a number of such scores were available. Although none could serve as a totally accurate representation of a cadet's performance, the combination of all measures provided a rough picture of such performance. Differences between the races obtained in peer ratings but not observed on other measures would be a possible indication that one of the subgroups was inappropriately disadvantaged by peer evaluations.

METHOD

The present investigation involved the use of data collected in 1976 at two of the three regional ROTC Advanced Camps. Only a nominal number of blacks attended the third camp, so data from that camp were not used. Subjects were 4545 ROTC cadets, including 3690 whites and 855 blacks. Various measures of cadet performance were administered at each camp by designated officials. These included two cadet peer evaluations, for which cadets rated the top ten and bottom ten individuals in their platoon on leadership potential and team member performance. An index reflecting the agreement of both blacks and whites with the overall platoon rating was also calculated at one of the camps. Cadets were also given two scores on overall camp performance and one score on performance in a tactical exercise by designated evaluators. Scores on a performance test designed to measure the cadet's ability to apply military skills, an orienteering test and a physical fitness test were also obtained.

RESULTS

Each of the questions examined in this study was explored through the use of multiple *t*-tests. As a precaution against the possibility of achieving a spurious finding of significance by this technique, the acceptable level of significance for each set of comparisons was reduced according to the number of *t*-tests performed. If the number of comparisons did not exceed 10, .005 was adopted as the appropriate level. If the number exceeded 10 but did not exceed 20, .0025 was the level used. It should also be noted that, although results are given for both leadership and team member peer ratings, the high correlation coefficient of .88 ($p < .001$) obtained between the two ratings across all 5598 cadets at the three Advanced Camps indicates that the two ratings are not really independent.

An examination of peer evaluation scores revealed that both blacks and whites favored their own subgroup. Mean scores given by blacks and whites to members of each of these two racial subgroups are shown in Table 1. These scores were examined in two types of comparisons. First, *t*-tests were used to compare each subgroup to subgroup rating with the score expected if both groups had performed at exactly the same level. The "expected" score was 2.00. Since 16 comparisons were involved, the significance level was set at .0025. All ratings were found to be significantly different from 2.00, with both blacks and whites consistently evaluating members of their own subgroup above this level and the other subgroup members below this level. For both the leadership and team member peer ratings, the highest subgroup to subgroup rating at each camp was that given by blacks to other blacks, followed by whites' ratings of other whites.

The second procedure used t-tests to compare ratings given within a particular racial subgroup with ratings given to that subgroup by the other. Eight comparisons were involved here, so the significance level was set at .005. At both camps, leadership and team member ratings given by whites to whites were significantly (p's $< .001$) higher than ratings given by blacks to whites. Similarly, leadership and team member ratings given by blacks to blacks were significantly (p's $< .001$) higher than ratings given by whites to blacks.

Given this evidence of same-subgroup favoritism by members of each race, the next set of analyses was directed at the question of how such favoritism was related to overall group disagreement. Within each platoon, an index of agreement between judgments of black raters and judgments of all raters was calculated on the basis of an accumulation of discrete comparisons between how individual blacks rated a particular cadet and how the entire platoon rated that cadet. On the basis of this index, platoons were classified as high or low in agreement. Differences between judgments rendered in high and low agreement platoons were analyzed by means of t-tests, with the results shown in Table 2. The significance level, based on the number of t-tests involved, was set at .005. Under the "Ratings to Blacks" section of the table, it can be seen that blacks assigned significantly higher same-subgroup ratings in the low agreement platoons than in the high agreement platoons. In the same section of Table 2, one finds that white ratings of blacks were just the reverse, being significantly higher in the high agreement platoons than in the low agreement platoons. A comparable pattern of divergence does not appear in the "Ratings to Whites" section of this table, where platoons are dichotomized according to agreement between white raters and all raters.

The peer rating results were then examined in the context of results on all Advanced Camp measures. Black and white scores on each measure, as shown in Table 3, were compared on the basis of t-tests. For six of these comparisons, where a preliminary test called into question the assumption of homogeneity of variance underlying the conventional t-test, a modified version of this test resulting in a more appropriate z statistic was used. The significance level for the total of 16 comparisons shown in this table was set at .0025. You will note a disparity between peer rating scores in Table 3 and those shown in Tables 1 and 2. The data in Table 3 are presented in terms of a score standardized to have an overall mean of 100, while Tables 1 and 2 present scores computed such that the overall mean is 2.00.

The t-test results indicated that black-white differences on peer ratings were not sharply divergent from differences between these subgroups on other Advanced Camp measures. If peer ratings are disregarded, whites received significantly higher scores than blacks on five out of six categories at Camp B and four of six at Camp A. If peer ratings are included, whites received significantly higher scores than blacks on seven of eight categories at Camp B and four of eight at Camp A.

DISCUSSION

Let us now consider the implications of the findings obtained here. This study essentially replicated the Mohr and Reidy (1976) finding of same-race favoritism in peer nominations in an ROTC Advanced Camp environment and provided evidence that the earlier finding was not merely an artifact of the particular rating instructions used in that study.

The comparison of ratings in platoons where black judgments were closely in accord with consensus judgments with ratings in platoons where such agreement was minimal provided further information about the rating patterns of each subgroup. As black ratings drifted further from consensus, black ratings of blacks became more favorable and white ratings of blacks became less favorable.

In examining possible explanations for this finding, one finds the concept of "race-bounded friendships" suggested in earlier studies a reasonable place to begin. A substantial body of literature (see Byrne, 1971) supports the proposition that individuals are attracted to those perceived as similar to themselves. Race presumably operates for many as at least a clue concerning the other's similarity. Thus, there may well be an initial predisposition to prefer same-race members in establishing friendships. This predisposition may be enhanced when the two major racial subgroups are represented unequally in the overall group. Such inequality may inhibit interracial friendships by emphasizing the distinctive nature of each subgroup, heightening the salience of subgroup membership and minimizing naturally occurring interactions between majority and minority members.

Any tendency by members of either subgroup to erect racial boundaries in the process of friendship formation obviously has implications for members of the other subgroup, who are likely to perceive this tendency and reciprocate. Furthermore, there is a basis for expecting race-bounded friendships to perpetuate themselves. Wilder and Allen (1978) have found that subjects choosing membership in one of two subgroups tend to prefer information which enhances their similarity to the chosen subgroup and their dissimilarity to the other. Perceived similarity with another is likely to influence attributions concerning the other's performance (Banks, 1976) and ultimately the evaluation of that performance. Thus, members of a race-bounded friendship subgroup might well be expected to rate one another more positively than they rate members of the other subgroup, an expectation reinforced by results from a number of studies showing a positive relationship between one's friendship with a peer and one's evaluation of that peer (Hollander, 1956; Hollander & Webb, 1955; Waters & Waters, 1970). Thus, race-bounded friendships may indeed produce the type of polarization between majority and minority subgroup member judgments observed in the present study, although the viability of this explanation relative to other hypotheses remains to be tested.

Despite the observed racial influences on peer evaluations in the present study, the relative performances of blacks and whites on these measures did not appear to depart substantially from the relative performances of these subgroups on other Advanced Camp measures. While the imperfect nature of these additional measures was commented on earlier, the consistency of racial differences across measures does suggest that peer ratings did not operate to the particular disadvantage of either racial subgroup.

Nevertheless, the evidence of racial influences on peer judgments is a matter of serious concern. This finding does not contradict the clear evidence from other studies that peer assessments are valid predictors of future performance, but does present a problem that needs to be addressed if such assessments are to be used to achieve their maximum potential.

Clearly, more research is needed to enable us to fully understand the processes that impact upon black and white peer evaluations. However, it is possible even at this point to identify approaches which appear promising as means of reducing racial influences upon these evaluations. Landy and Farr (1980) have suggested that rating errors can be reduced if rating formats incorporating behavioral anchors are used, although the expected gains from this procedure are relatively modest. Rater training may also be beneficial. Such training might incorporate such topics as: a description and discussion of the rating format used, a discussion of the importance of behavior as a basis for evaluation, guidance on how to observe behavior, a discussion of racial stereotypes and their contribution to rating errors, and a discussion of how to avoid rating errors associated with selective attention and inaccurate attributions.

If race-bounded friendships are indeed primarily responsible for the results observed in this study, then approaches which fail to deal directly with such friendships are likely to have limited impact. Perhaps the most effective mechanism for reducing racial influences, where feasible, would be to modify the racial composition of the units in which ratings are given. A distribution of cadets such that, in those units in which blacks are represented, they are represented in equal numbers with white cadets, might well create an environment less conducive to the formation of race-bounded friendships than the one examined here. To the extent that racial boundaries in the development of friendship groups interfere with free interactions between blacks and whites, one might well consider such an environmental change beneficial quite apart from its anticipated positive effect on rating accuracy.

REFERENCES

Banks, W. C. The effects of perceived similarity upon the use of reward and punishment. Journal of Experimental Social Psychology, 1976, 12, 131-138.

Byrne, D. The attraction paradigm. New York: Academic Press, 1971.

Clore, G. L., Bray, R. M., Itkin, S. M., & Murphy, P. Interracial attitudes and behavior at a summer camp. Journal of Personality and Social Psychology, 1978, 36, 107-116.

Cox, J. A., & Krumboltz, J. D. Racial bias in peer ratings of basic airmen. Sociometry, 1958, 21, 292-299.

DeJung, J. E., & Kaplan, H. Some differential effects of race of rater and ratee on early peer ratings of combat aptitude. Journal of Applied Psychology, 1962, 46, 370-374.

Downey, R. G., Medland, F. F., & Yates, L. G. Evaluation of a peer rating system for predicting subsequent promotion of senior military officers. Journal of Applied Psychology, 1976, 61, 206-209.

Haggerty, H. R. Status report on research for the U. S. Military Academy (cadet leaders task). (Army Research Institute Technical Research Report 1133) Alexandria, Va.: Army Research Institute, 1963.

Hollander, E. P. The friendship factor in peer nominations. Personnel Psychology, 1956, 9, 435-447.

Hollander, E. P., & Webb, W. B. Leadership, followership, and friendship: An analysis of peer nominations. Journal of Abnormal and Social Psychology, 1955, 50, 163-167.

Kane, J. S., & Lawler, E. E. III. Methods of peer assessment. Psychological Bulletin, 1978, 85, 555-586.

Landy, F. J., & Farr, J. L. Performance rating. Psychological Bulletin, 1980, 87, 72-107.

Lewin, A. Y., & Zwany, A. Peer nominations: A model, literature critique and a paradigm for research. Personnel Psychology, 1976, 29, 423-447.

Mohr, E. S., & Reidy, R. F. Racial bias in peer ratings at ROTC Advanced Summer Camp, Fort Bragg, 1975. (Army Research Institute Research Memorandum 76-22) Alexandria, Va.: Army Research Institute, 1976.

Rumsey, M. G. Gender influences on peer ratings in ROTC training platoons. Paper presented at Annual Meeting of the American Psychological Association, Los Angeles, August, 1981.

Schmidt, F. L., & Johnson, R. H. Effect of race on peer ratings in an industrial situation. Journal of Applied Psychology, 1971, 57, 237-241.

Waters, L. K., & Waters, C. W. Peer nominations as predictors of short-term sales performance. Journal of Applied Psychology, 1970, 54, 42-44.

Wilder, D. A., & Allen, V. L. Group membership and preference for information about others. Personality and Social Psychology Bulletin, 1978, 4, 106-110.

ACKNOWLEDGEMENT

The creative and knowledgeable statistical analysis and consulting assistance provided by Mrs. Gail Rowan is gratefully acknowledged by the author.

TABLE 1

Peer Rating Scores by and for Blacks and Whites

		Ratings Given by Blacks		Ratings Given by Whites		
Rated Group	N	M*	SD	M*	SD	t**
		Camp A: Leadership Rating				
Blacks	444	2.39	.36	1.87	.37	20.08
Whites	1740	1.91	.33	2.04	.32	11.18
		Camp A: Team Member Rating				
Blacks	444	2.42	.39	1.88	.39	20.88
Whites	1740	1.91	.37	2.03	.38	10.08
		Camp B: Leadership Rating				
Blacks	411	2.48	.39	1.82	.38	24.15
Whites	1950	1.92	.38	2.04	.40	9.92
		Camp B: Team Member Rating				
Blacks	411	2.39	.38	1.79	.36	23.00
Whites	1950	1.93	.36	2.05	.33	10.45

*All t values comparing means with 2.00 significant at $p < .0025$
**All t values shown significant at $p < .001$

TABLE 2

Ratings Received by Whites and Blacks
in Platoons with High and Low Agreement Indices

	High Agreement			Low Agreement			
Rating Type	N	M	SD	N	M	SD	t
			Ratings to Whites				
Black Ldr	23	1.93	.03	22	1.89	.04	3.75*
White Ldr	23	2.04	.04	22	2.03	.03	1.00
Black TM	21	1.92	.04	23	1.89	.05	2.62
White TM	21	2.03	.04	23	2.04	.04	.66
			Ratings to Blacks				
Black Ldr	24	2.31	.13	24	2.50	.16	4.49*
White Ldr	24	1.93	.13	24	1.79	.10	4.18*
Black TM	24	2.35	.26	24	2.53	.20	3.27*
White TM	24	1.94	.13	24	1.81	.09	3.96*

*$p < .005$

TABLE 3

Scores Received by Blacks and Whites
on all Advanced Camp Measures

Variables	Whites N	Whites M	Whites SD	Blacks N	Blacks M	Blacks SD	t
			Camp A				
SOAT	1734	100.37	5.83	443	98.85	6.43	4.47*[a]
PNAT	1740	100.34	19.40	444	99.05	20.64	1.24
POAT	1740	100.84	19.54	444	97.39	19.78	3.32*
PEER (TM)	1740	100.59	19.02	444	97.82	21.57	2.47[a]
PEER (LDR)	1740	100.36	19.51	444	99.04	19.94	1.21
PT	1734	407.39	39.50	444	424.06	37.75	8.01*
ORIENT	1740	101.99	15.40	444	92.92	16.72	10.93*
MIL STAKES	1738	102.91	18.31	444	88.53	20.38	13.57*[a]
			Camp B				
SOAT	1934	100.36	5.83	408	99.15	6.38	3.56*[a]
PNAT	1950	101.25	19.39	411	95.83	20.02	5.11*
POAT	1950	101.25	19.56	411	95.55	18.73	5.43*
PEER (TM)	1950	101.37	19.25	411	93.16	19.53	7.82*
PEER (LDR)	1950	100.82	19.59	411	96.08	18.67	4.51*
PT	1948	429.02	41.06	411	445.31	32.84	8.71*[a]
ORIENT	1950	101.79	15.34	411	93.01	15.47	10.58*
MIL STAKES	1950	101.81	18.58	411	91.30	21.32	9.30*[a]

*$p < .0025$
[a] z statistic

Brief Description of Variables Presented in Table 3

SOAT: Rating of cadet performance in a one-day tactical exercise.

PNAT: Rating of cadet overall camp performance by a non-commissioned officer.

POAT: Rating of cadet overall camp performance by an officer

PEER (TM): Peer rating on team member performance.

PEER (LDR): Peer rating on leadership potential.

PT: Performance on a physical fitness test. Here, total raw score, rather than a standardized score with a mean of 100, is used.

ORIENT: Score on a timed freestyle orienteering performance test.

MIL STAKES: Score on a performance test designed to measure the cadet's ability to apply military skills.